

A Novel Ray Analogy for Enrolment of Ear Biometrics

Alastair H. Cummings, Mark S. Nixon, John N. Carter

Abstract—The ear is a maturing biometric with qualities that give it superiority over other biometrics in a number of situations; in particular the ear is relatively immune to variation due to ageing. Successful ear biometrics rely upon a well enrolled dataset, with ears normalised for position, scale and rotation. We present a novel ear enrolment technique using the image ray transform, based upon an analogy to light rays. The transform is capable of highlighting tubular structures such as the helix of the ear and spectacle frames and, by exploiting the elliptical shape of the helix, can be used as the basis of a method for enrolment for ear biometrics. The presented technique achieves 99.6% success at enrolment across 252 images of the XM2VTS database, displaying a resistance to confusion with hair and spectacles. These results show great potential for enhancing many other already existing enrolment methods through use of the image ray transform at a preprocessing stage.

I. INTRODUCTION

Ears are a promising avenue for biometrics research and have been found to be able to identify subjects with similar performance to face recognition [1]. The ear biometric problem (as with all biometrics) is split between enrolment and recognition. Enrolment is the discovery, localisation and normalisation of the ear image, whilst recognition deals with identification of a subject by ears. Currently, most enrolment for experiments is done manually, but in any real world test of ear biometrics, enrolment is just as important a problem as recognition. Enrolment has not seen the same amount of research as recognition but there is some development in the area. Alvarez [2] used active contours and an ovoid model to find the ear, and in a similar vein Arbab-Zavar and Nixon [3] used an elliptical Hough transform. Bustard and Nixon [4] developed a method of transforming SIFT feature points from an image of an ear to a model, and then used the distance to create a biometric that was resilient to pose variation and occlusion. Islam et al. [5] used weak classifiers based upon Haar wavelets combined using AdaBoost to create a very successful ear detection method. Prakash et al. [6] used both colour and shape to achieve 95.2% enrolment success in a database of 150 side view images, suggesting that colour can indeed augment the enrolment process.

Physical analogies are an exciting paradigm in computer vision that enable the creation of novel techniques that approach the problems of feature extraction from entirely different angles [7]. These analogy based techniques have the advantage of being based on physical properties of natural phenomena such as water, heat or force fields and so are

more easily understood by those using them. In addition to the intuitive nature of the algorithms, the parameters that are used have meanings that are clear and have real world analogues. Although analogy operators are heavily based upon a well defined physical concept, the analogies can be adapted outside this definition to increase their effectiveness and flexibility whilst maintaining the strengths provided by the analogy. These properties are a clear advantage over many standard techniques for which the mechanics can be hard to grasp and parameter selection is not clear.

Heat flow has been used as an analogy due to its smoothing properties. Anisotropic diffusion [8] is an edge-aware smoothing technique that allows heat to flow across areas of low but not high edge strength allowing Gaussian noise to be eliminated whilst maintaining important edge features. This is one of the earliest examples of a principled vision technique based upon an analogy.

Hurley's force field transform [9] generates a force field from an image that is analogous to a gravitational or magnetic field. Each pixel is assumed to attract every other pixel with a force dependent on its intensity and the inverse square law. The sum of these forces generates a vector field representing an image. This force field can help in feature extraction and was successfully used to create an ear biometric. Xie and Mirmehdi [10] also used a force field analogy, generated from magnetostatic theory and joined with an active contour model to enable contour detection. In their model, the image border and evolving contour are assumed to have an electric current running through them and the interaction of these currents generates a force field. This field guides the development of the contour, changing along with it to guide it to the image border.

We use the image ray transform, a transform based upon an analogy to light rays, to enable ear enrolment. Previously it has been used to enhance circle detection [11], but we take advantage of its ability to highlight tubular features. Other techniques have not used light in such a strongly analogical sense; however, the Eikonal equation has been used many times for vision techniques. The Eikonal equation describes the time a ray takes to travel from the boundary of an anisotropic medium to a point within it and has been used in a number of applications [12], none of which fully take advantage of the possible analogical formulation and it is most often used as a distance metric.

This paper describes an ear enrolment technique using the image ray transform. Section II describes the image ray transform and its properties. Section III describes work into applying the transform for ear biometric enrolment, showing how its properties are well matched to the ear's

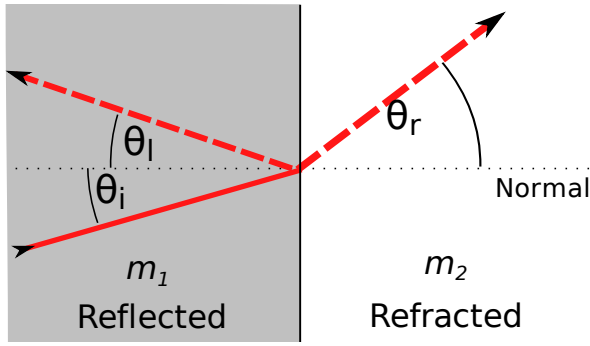


Fig. 1. Refraction and reflection of light.

structure. Finally section IV draws conclusions and describes future work: both expanding on the applications previously described and others that have yet to be investigated.

II. THE IMAGE RAY TRANSFORM

The image ray transform is a novel technique for extracting tubular and circular features that are not often found by other methods. It uses a pixel based ray tracing technique and a subset of the laws of optics to trace rays through an image. These then react to certain structural features to emphasise them. Whilst the transform is based on the principles of optics, the details of the technique can be adjusted to suit successful feature extraction rather than accurate simulation of the natural phenomenon. The implementation capitalises only on the basis of the analogy; we do not simulate intricate details of light propagation.

A. Laws of Optics

Rays are a method of modelling the propagation of waves, most often light, with specific regard for the direction of the wave as it interacts with its environment and ignoring wave-like interactions such as diffraction. The path of a light ray will be altered when it crosses the boundary with a medium of different refractive index, refracting and/or reflecting (see figure 1). Light crosses the boundary between the media at an angle of θ_i between the light direction and the normal of the boundary (the dotted line). If reflection occurs, then $\theta_r = \theta_i$. If refraction occurs, the light refracts at an angle of θ_r to the normal where θ_r is calculated from the refractive indices of n_1 and n_2 of the media m_1 and m_2 . Refractive indices are the ratio of the speed of light in a vacuum to the speed of light within the media, and in nature $1 \leq n \lesssim 4$. The angle of the refracted light θ_r is found by Snell's Law as

$$\frac{\sin \theta_i}{\sin \theta_r} = \frac{n_2}{n_1} \quad (1)$$

If $n_1 < n_2$, then light bends towards the normal, so $\theta_r < \theta_i$. If $n_1 > n_2$, then light bends away from the normal, as long as it is below the critical angle θ_c . This is the angle for which θ_r would be greater than 90° and is therefore physically impossible and is calculated as

$$\theta_c = \sin^{-1} \left(\frac{n_2}{n_1} \right) \quad (2)$$

In this case, the light is totally internally reflected in an identical manner to normal reflection. In the natural world, the amount of refraction and reflection that occurs depends on the media and in most cases is a combination of the two, some part passing through, and some being reflected back. These rules form the basis of the image ray transform.

B. The Image Ray Transform

The image ray transform takes the laws of optics described previously and applies them to the problem of feature extraction. The transform operates by tracing a number of rays through an image. The paths of these rays are then used to generate a new image, with tubular and circular features emphasised. The image is analogised to a matrix of two dimensional glass blocks, each representing a pixel whose refractive index is related to the intensity of the pixel in the image. One method to assign a refractive index to a pixel with intensity i is shown in equation 3, where n_{\max} defines the maximum possible refractive index. The indices are then spaced evenly between 1 and n_{\max} .

$$n_i = 1 + \left(\frac{i}{255} \right) \cdot (n_{\max} - 1) \quad (3)$$

The splitting of rays into reflected and refracted parts is not considered in order to reduce computational complexity.

In this matrix of blocks, a ray is created with the position (x and y) and the direction (ϕ) drawn from a uniform distribution (U). For an image of size $w \times h$,

$$x \sim U[0, w), y \sim U[0, h), \phi \sim U[0, 2\pi) \quad (4)$$

ϕ is converted into a unit vector, \mathbf{V} , representing the direction of the ray (this is more convenient for calculations, but cannot be directly generated from a uniform distribution):

$$\mathbf{V} = (\cos \phi, \sin \phi)^T \quad (5)$$

The ray function r traces the course of the ray and updates an accumulator \mathbf{A} . \mathbf{I} is the image from which refractive indices and the normals at each pixel are derived. d is the maximum number of reflections or refractions which the ray can undergo before tracing ceases, also known as the depth.

$$\mathbf{A}' = r(\mathbf{A}, \mathbf{I}, x, y, \mathbf{V}, d, n_{\max}) \quad (6)$$

The position vector \mathbf{p} initialises the ray at:

$$\mathbf{p}^{<0>} = (x, y)^T \quad (7)$$

and at iteration i :

$$\mathbf{p}^{<i+1>} = \mathbf{p}^{<i>} + \mathbf{V} \quad (8)$$

At each pixel through which the ray passes, \mathbf{A} is updated to show that the ray has passed through:

$$\mathbf{A}'(\mathbf{p}^{<i+1>}) = \mathbf{A}(\mathbf{p}^{<i+1>}) + 1 \quad (9)$$

$\mathbf{A}(\mathbf{p})$ is only increased once per ray, so as to prevent small loops in the ray's path repeatedly increasing a single pixel and excessively emphasising noise or other undesirable features. When crossing a pixel boundary between media of differing refractive indices, a new path must be calculated

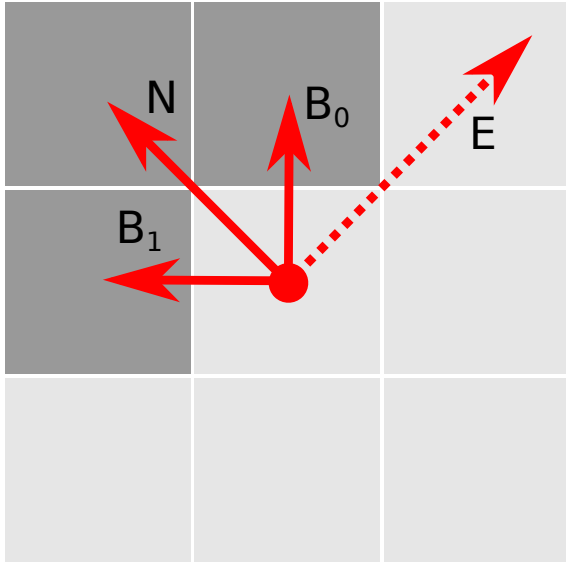


Fig. 2. The normals that could be used to calculate reflections and refractions.

using a vector formation of the laws described in section II-A.

If \mathbf{N} is the normal and n_1 and n_2 are the refractive indices of the first and second media respectively, then the angle of incidence θ_i , can be found:

$$\cos \theta_i = \mathbf{N} \cdot \mathbf{V} \quad (10)$$

If $n_1 > n_2$, we have to test whether we should internally reflect or refract. The critical angle θ_c is found from equation 2 and if $n_2 > n_1$ or $\theta_i < \theta_c$ and assigning $n = \frac{n_1}{n_2}$ then the direction of refraction \mathbf{R}_r is, from [13]:

$$\mathbf{R}_r = n\mathbf{V} + (n(\mathbf{N} \cdot \mathbf{V}) - \cos \theta_r)\mathbf{N} \quad (11)$$

where $\cos \theta_r$ is

$$\cos \theta_r = \sqrt{1 - n^2(1 - \mathbf{N} \cdot \mathbf{V})} \quad (12)$$

If $n_1 > n_2$ and $\theta_i > \theta_c$, then the ray totally internally reflects. The direction of reflection \mathbf{R}_l is then

$$\mathbf{R}_l = \mathbf{V} - 2(\mathbf{N} \cdot \mathbf{V})\mathbf{N} \quad (13)$$

\mathbf{R}_l or \mathbf{R}_r are then assigned to \mathbf{V}

$$\mathbf{V}' = \begin{cases} \mathbf{R}_l & \text{if } n_1 > n_2 \wedge \theta_i > \theta_c \\ \mathbf{R}_r & \text{otherwise} \end{cases} \quad (14)$$

The ray is traced in this manner until it has either undergone d reflections or refractions, or the ray exits the image. This is repeated for N rays or until a stopping condition (section II-C.2) is reached, and the normalised accumulator gives the transformed image. The values of the normals (\mathbf{N}) used to calculate the new directions of rays are always set to the normal of the edge direction found by the Sobel operator at that point. In figure 2, this normal is shown orthogonal to the edge direction \mathbf{E} , as well as the normals of the pixel boundaries \mathbf{B}_1 and \mathbf{B}_2 . \mathbf{N} is used rather than \mathbf{B}_1 or \mathbf{B}_2 as it is more representative of the information within the image.

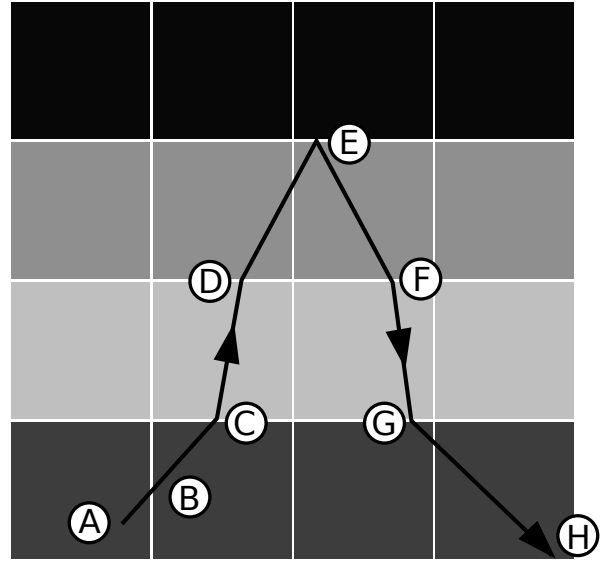


Fig. 3. An example of the course a ray might take in a simple 4x4 image.

Currently, we only perform refraction when moving from higher to lower refractive indices as this tends to improve the quality of the transform result, and we are working to improve this further.

Figure 3 shows the path of a ray as it travels through a simple image. The edge direction at each pixel is horizontal, and hence the normal at each point is vertical. The ray is initialised at position A with a random direction and advances to position B. The intensity of the pixels indicates their refractive index. At B, the refractive indices of the current and next pixel are compared (n_1 and n_2 respectively), and as $n_1 = n_2$ the ray continues with no change in direction. At C, $n_1 < n_2$, and so the ray bends towards the normal, whereas at D $n_1 > n_2$ so the ray bends away from the normal. At E, total internal reflection occurs, because $n_1 > n_2$ and $\theta_i > \theta_c$. At F and G, the occurrences at C and D are repeated in reverse order, whilst at H the ray exits the image and the trace terminates.

C. Refinements

In addition to the basic transform, a number of additional parameters and processes can be performed with the transform to give a wider range of results and to extend the range of circumstances in which the transform is useful.

1) *Intensity and Refractive Index:* The transform will only extract features that have larger refractive indices than the surrounding area because total internal reflection only prevents rays from passing from materials with higher to lower refractive indices, and not vice versa. As the refractive index is proportional to intensity, this has the effect of extracting features that are lighter than their surrounding area. Depending on what is known about the properties of the features to be extracted different measures can be taken to improve the result of the transform.

If the approximate intensity of the desired features is known then the image can be easily transformed to make

the target intensity have the largest value. This can be done by finding the difference from the target intensity t to the original intensity i_o for each pixel, as in equation 15.

$$i_t = |i_o - t| \quad (15)$$

If the expected intensity is not known then the ray transform can be performed multiple times, with different target intensities, and the largest value at each pixel used. This is shown in equation 16 where \mathbf{I}' is the final image and \mathbf{I}_x are the results of different transforms with different target intensities.

$$\mathbf{I}' = \max(\mathbf{I}_0, \mathbf{I}_1, \dots, \mathbf{I}_x) \quad (16)$$

2) *Automatic Stopping Condition:* Finding the optimal number of rays that should be traced in order to ensure the result of the transform is of sufficient quality is a challenge that must be met. Rather than use fixed values of N , an improved method for automatically deciding when the transform should cease is to monitor the resultant image and stop when it no longer changes significantly between iterations.

The best method that has been found to do this is to measure the difference in the normalised accumulator image between iterations. We use the root mean squared (RMS) difference between the intensities:

$$D^{<t>}(\mathbf{I}'^{<t>}, T) = \sqrt{\frac{1}{|\mathbf{I}'|} \sum_{i \in \mathbf{I}'} \left(\mathbf{I}'^{<t>}(i) - \mathbf{I}'^{<t-T>}(i) \right)^2} \quad (17)$$

where T is the number of iterations between each comparison. Rather than using successive images, the operation of a transform is sufficiently fast that large intervals should be used, that is, $T \gg 1$. The results produced by this method are consistent with the observed resultant images, with the size of D accurately reflecting the change that can be observed in the successive images.

The RMS difference measure is also cheap to calculate when T is set high enough, as expensive accumulator normalisations are calculated infrequently. It can be used as a stopping condition by setting a minimum value (D_S) to stop the process. From experimentation it has been found that a value for D_S of 1 when $T = 1000$ allows termination with a stable result.

D. Properties of the Transform

The ray transform is capable of emphasising tubular features within an image. It is able to do this because tubular features in the image which have larger refractive indices than the surrounding area can act like waveguides (such as an optical fibre), guiding rays along the length of the structure via total internal reflection. Rays that enter tubes at shallow angles are likely to reflect along the length and concentrate rays inside it. Evidence of this can be seen in figure 4. The transform is also adept at highlighting circles through repeated shallow reflections off the edges, and this has previously been used to enhance circle detection with the Hough Transform [11]. Figure 4(b) shows a single ray as it passes through an image with the original image faintly in

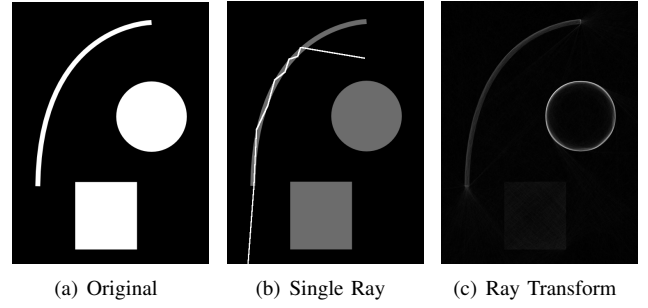


Fig. 4. The application of the ray transform to a 480x640 image with a tubular feature. For 4(c) $N = 10000$, $d = 256$ and $n_{\max} = 2$.

the background. The ray is initialized above the circle, and travels towards the curve (approximately 10 pixels wide), entering it. On reaching the far edge of the curve, it reflects, as the angle of incidence is above the critical angle. This continues until it exits the curve and, subsequently, the image. The final image can be seen in figure 4(c), where the curve and the circle are more heavily emphasised than the box.

Tubular structures are found in many real images. One example that clearly shows the ray transform's strength at extracting tubular features is the helix of the human ear. The helix is the outer section of the ear, which appears from a profile view as a curved, tubular structure. By applying the ray transform to such images, it is possible to extract the helix of the ear and aid enrolment for ear biometrics.

III. EAR ENROLMENT WITH THE IMAGE RAY TRANSFORM

The image ray transform's propensity for highlighting tubular features such as the helix of the ear can be exploited to create an enrolment technique. This technique uses the transform to create an image in which enrolment is considerably easier than in any edge or intensity image. Enrolment then occurs using a simple template matching technique, but it should be noted that many other enrolment techniques could be enhanced through use of the image ray transform as a preprocessing step on the original ear image.

A. Enrolment Technique

The technique used here exploits the image ray transform's strength at extracting tubular features to highlight the helix of the ear and then uses a series of simple steps to extract and normalise the ear. It should be noted that the main contribution of this work is the application of the image ray transform: without doubt more complex techniques (including some of the alternative enrolment methods described above) could be used to increase the success of the enrolment process.

The initial step is to apply the image ray transform to the ear image with the parameters $D_S = 1$, $n_{\max} = 40$, $d = 256$. These parameters are standard values that produce an image that has had enough rays of sufficient length traced to reduce the noise to an manageable level, and the high value of n_{\max}

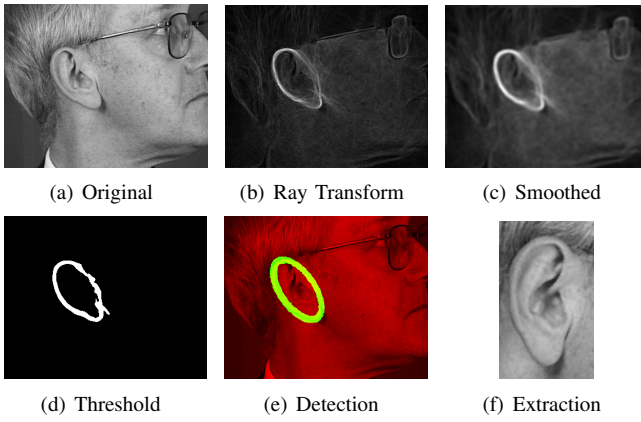


Fig. 5. Example of the steps taken to achieve successful ear enrolment.

makes rays conform strongly to structural features within the image. This produces an image in which the helix of the ear is highlighted (figure 5(b)) (in addition to other tubular features such as spectacles). Due to the non-deterministic nature of the transform, it is then necessary to apply Gaussian smoothing to the result (figure 5(c)) to reduce noise. The image is then thresholded (figure 5(d)) to produce an image with a strong helix as its focus. Histogram equalisation is used to allow a consistent threshold (of 249) across all images to be used. Simple template matching is then used with an elliptical template across a range of rotations and scales (figure 5(e)) and then the matched section is then normalised and extracted (figure 5(f)).

B. Results

Enrolment was performed on 252 images (4 per subject) from the XM2VTS database [14] using the technique above. This is a section of the database that has been used earlier, and provides basis for comparison. The mean computation time for each ear for the entire technique was 5.45s, whilst the ray transform alone took only 1.47s and required approximately 19000 rays to be traced. In figure 6, a range of images after the thresholding stage are shown. The propensity of the technique to extract the helix strongly in all images is shown, as well as cases where spectacle frames (figures 6(a),(b) and (c)) or other features are highlighted (light hair in figure 6(a) and part of the forehead in figure 6(d)). The extent to which features are highlighted depends strongly upon their intensity (as discussed in section II-C.1), primarily highlighting ear helices due to their high intensity skin surrounded by low intensity shadows. Other structures such as some spectacle frames, and, more rarely, hair, are highlighted through this mechanism as well. In general it was found that these extra features retained through smoothing and thresholding do not affect the enrolment results in most cases, because they are rarely of an elliptical shape of the type that would be found in the template matching stage.

The results of the extraction were encouraging. Out of 252 images, 99.6% of extracted images contained the subject ear. 98.4% of these images had the ear correctly normalised,

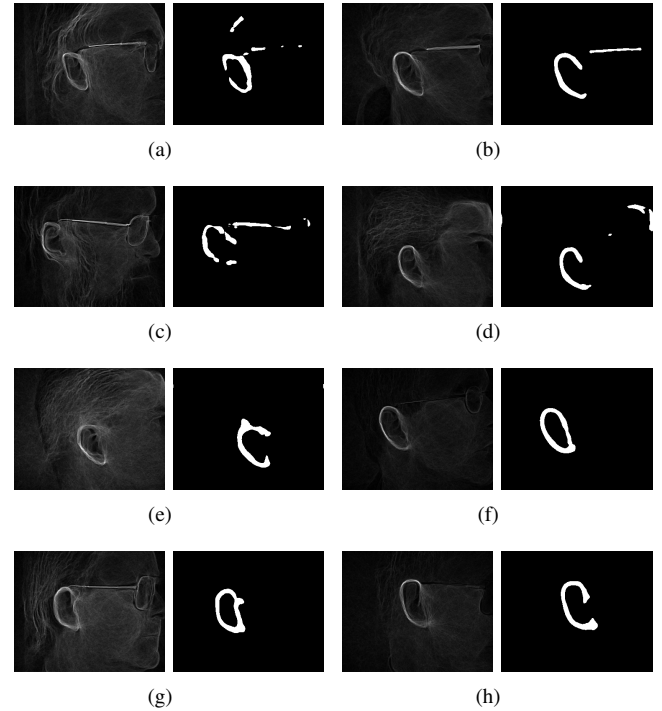


Fig. 6. Selection of transformed images before and after smoothing and thresholding.

with the incorrectly normalised images having small scale or rotation errors. This is a performance similar to that achieved earlier [3]. It is possible that the enrolment process is more refined by our new approach, but certainly it appears that the preprocessing here could be of advantage in the other technique, which used edge detection.

The successful performance of the technique and small number of failed enrolments has prevented meaningful ROC analysis. Further experiments in which peaks which were not the maximal value in images were extracted and classified would allow the discriminatory power of the technique to be more accurately characterised.

Shown in figure 8 is the image in which ear extraction failed completely. The ear was extracted perfectly in all other images of this subject. The ear has not been extracted due to both a streak of white hair and the subject's spectacle frames being strongly extracted in addition to the ear. The combination of these features has created a structure that responds more strongly in the template matching phase than the ear. This situation could be successfully prevented by filtering out straight features that are very unlikely to be part of the ear, such as the spectacle frames, before matching occurs.

IV. DISCUSSION AND FUTURE WORK

We have shown how the image ray transform is a novel and powerful technique that can be used to enhance detection of structural features. In particular, the transform's aptitude for highlighting tubular features has been employed in order to create an effective technique for ear enrolment. Further,

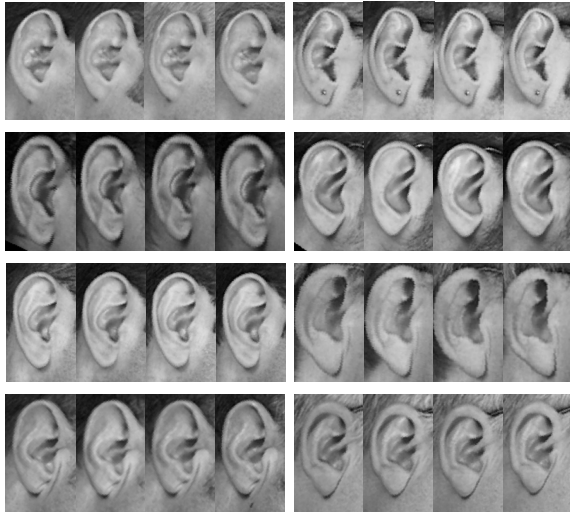


Fig. 7. Extracted and normalised ears for a selection of subjects



Fig. 8. The ear image that failed to be extracted correctly

the computational complexity is such that it appears to be an interesting contender for a choice as a preprocessing stage. The transform has an inherent ability to strongly highlight the helix of the ear in all cases, and only rarely highlights other facial features in similar ways, making extraction simple and reliable. Clearly, we have a low-level feature extraction technique, which can be specifically tuned to ear extraction. The technique shown here concentrates on exhibiting the potential of the image ray transform for enhancing enrolment and relies only upon simple thresholding and template matching. It is likely that through use of a more complex elliptical Hough transform, such as the one utilised by Arbab-Zavar and Nixon [3], these results could be improved upon further.

There are many directions in which the image ray transform and its application to ear biometrics could be expanded. The transform itself could be improved with more

targeted ray initialisations; drawing initial ray positions from a probability distribution with peaks at likely ear positions. Further work could be done on enhancing enrolment with an improved matching algorithm, or, alternatively, creating a recognition technique based upon the transformed ear. The ray transform may also be appropriate for use in gait biometrics, as legs act as tubular features that the transform is adept at extracting. The transform could also be extended to work upon 3D images, both spatial and spatio-temporal, for 3D biometrics or object tracking.

ACKNOWLEDGEMENTS

We gratefully acknowledge Alastair Cummings' EPSRC CASE studentship funded by the National Physical Laboratory (NPL).

REFERENCES

- [1] K. Chang, K. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1160–1165, September 2003.
- [2] L. Alvarez, E. Gonzalez, and L. Mazorra, "Fitting ear contour using an ovoid model," in *Proc. 39th Annual Int'l Carnahan Conf. Security Technology, (CCST'05)*, 2005, pp. 1–4.
- [3] B. Arbab-Zavar and M. Nixon, "On shape-mediated enrolment in ear biometrics," in *Proc. Int'l Symposium Visual Computing (ISVC07)*, November 2007.
- [4] J. D. Bustard and M. Nixon, "Robust 2d ear registration and recognition based on sift point matching," in *Proc. IEEE 2nd Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS08)*, September 2008.
- [5] S. Islam, M. Bennamoun, and R. Davies, "Fast and fully automatic ear detection using cascaded adaboost," in *Proc. IEEE Workshop on Application of Computer Vision (WACV 2008)*, 2008, pp. 1–6.
- [6] S. Prakash, U. Jayaraman, and P. Gupta, "A skin-color and template based technique for automatic ear detection," in *Proc. 7th Int'l Conf. Advances in Pattern Recognition. (ICAPR'09)*, 2009, pp. 213–216.
- [7] M. S. Nixon, X. U. Liu, C. Direkoglu, and D. J. Hurley, "On using physical analogies for feature and shape extraction in computer vision," *The Computer Journal*, 2009.
- [8] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [9] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Computer Vision and Image Understanding*, vol. 98, pp. 491–512, November 2005.
- [10] X. Xie and M. Mirmehdi, "MAC: Magnetostatic active contour model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 632–647, 2008.
- [11] A. H. Cummings, M. S. Nixon, and J. N. Carter, "Circle detection using the image ray transform," in *Int'l Conf. Computer Vision Theory and Applications (VISAPP 2010)*, 2010.
- [12] P. Maragos, "PDEs for morphological scale-spaces and eikonal applications," in *The Image and Video Processing Handbook*, 2nd ed., A. C. Bovik, Ed. Elsevier Academic Press, 2005, ch. 4.16, pp. 587–612.
- [13] F. Hill, *Computer graphics using OpenGL*, 3rd ed. Prentice Hall, 2000, ch. 12, p. 678.
- [14] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int'l Conf. Audio and Video-based Biometric Person Authentication*, 1999, pp. 965–966.